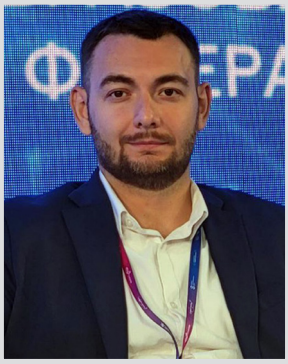
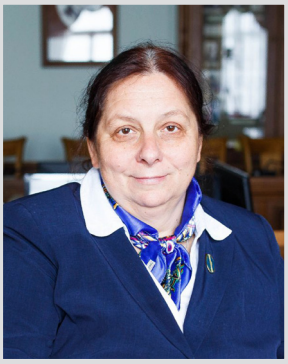
УДК 34
DOI: 10.24412/1998-5533-2025-3-235-241

Преступления, совершаемые с использованием искусственного интеллекта: анализ рисков и правовые перспективы



Талан А.С.
Кандидат химических наук,
заместитель генерального директора
ООО «Актру-Образовательные технологии» (Химки)



Талан М.В.
Доктор юридических наук, профессор,
заведующая кафедрой уголовного права
Казанского (Приволжского) федерального университета,
профессор кафедры уголовного права и процесса
Университета управления «ТИСБИ» (Казань)

Отсутствие систематизированного подхода к классификации правонарушений с использованием искусственного интеллекта создает серьёзные пробелы в правовом регулировании и правоприменительной практике.

Цель исследования заключается в разработке комплексной классификации преступлений, совершаемых с применением технологий искусственного интеллекта, а также формулировании предложений по совершенствованию регулирования данной сферы.

Задачи работы включают: систематизацию существующих и потенциальных видов ИИ-преступлений; анализ применимости действующих норм российского уголовного права к новым видам правонарушений; выявление правовых пробелов в регулировании ИИ-преступности; разработку конкретных законодательных инициатив.

Научная значимость исследования определяется впервые предложенной пятикомпонентной классификацией преступлений, связанных с использованием искусственного интеллекта, которая охватывает спектр угроз от технических уязвимостей до психологических манипуляций.

Практическая ценность работы состоит в разработке конкретных предложений по внесению изменений в ст.ст. 272, 273 УК РФ и введению новой ст. 273.1 УК РФ, а также в обосновании системы квалифицирующих признаков, специфичных для ИИ-преступлений.

Основные результаты исследования представлены пятью категориями ИИ-преступлений: отравление данных языковых моделей, вирусные атаки на ИИ-системы, перехват данных ИИ-агентов, захват управления ИИ или ИИ-агентов и создание сгенерированного ИИ контента для социальной инженерии. Проведённый системный анализ уязвимостей ИИ-агентов в контексте уголовно-правовой доктрины выявил правовые пробелы в действующих нормах главы 28 УК РФ. Выводы работы обосновывают необходимость законодательных изменений, которые включают усиление санкций с учетом потенциального ущерба от ИИ-преступлений и создание специализированного правового механизма противодействия новым видам киберугроз.

Ключевые слова: *искусственный интеллект, киберпреступность, уголовное право, ИИ-агенты, классификация преступлений*

Для цитирования: *Талан А.С., Талан М.В. Преступления, совершаемые с использованием искусственного интеллекта: анализ рисков и правовые перспективы // Вестник экономики, права и социологии. 2025. № 3. С. 235–241. DOI: 10.24412/1998-5533-2025-3-235-241.*

Современное развитие технологий искусственного интеллекта создаёт не только новые возможности для общества, но и порождает принципиально новые виды преступлений. По данным МВД России, в 2024 г. количество преступлений с использованием информационно-телекоммуникационных технологий достигло 765,4 тыс. случаев, что составляет рекордные 40 % от общего числа зарегистрированных преступлений – исторический максимум [1]. Это представляет рост на 13,1 % по сравнению с предыдущим годом, при том что доля ИТ-преступлений неуклонно растет с 25 % в 2020 г. [2].

Особую тревогу вызывает отсутствие адекватного правового регулирования в сфере ИИ-преступности. Как отмечают исследователи, «использование искусственного интеллекта в целях реализации преступного умысла не имеет отражения в Уголовном кодексе Российской Федерации» [3]. В то же время более 90 % российских компаний уже используют большие языковые модели в своей деятельности, а около 70 % организаций сталкивались с атаками через ИИ-системы [4].

ИИ-агенты представляют собой качественно новый класс программных систем, существенно отличающийся от традиционных ИИ-решений. Согласно современным определениям, интеллектуальный агент – это «программа, самостоятельно выполняющая задание, указанное пользователем компьютера, в течение длительных промежутков времени» [5]. Ключевые характеристики ИИ-агентов включают [6]:

• **автономность**: способность действовать без постоянного контроля человека;

• **реактивность**: восприятие изменений окружающей среды и адекватное реагирование;

• **проактивность**: инициирование действий для достижения поставленных целей;

• **социальность**: взаимодействие с другими агентами и людьми.

В отличие от традиционных ИИ-систем, работающих по принципу «запрос – ответ», агенты способны **планировать многошаговые операции, адаптироваться к изменяющимся условиям и самостоятельно принимать решения** [7]. Национальная стратегия развития ИИ определяет искусственный интеллект как «комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека» [8].

Современные исследования выявляют **критические уязвимости ИИ-агентов**, создающие новые векторы атак [9]:

Prompt Injection – наиболее распространенный тип атак, составляющий свыше 40 % всех атак на ИИ-системы [10]. Злоумышленники внедряют вредоносные инструкции в запросы к модели, заставляя ее действовать непредусмотренным образом.

Пример базовой *Prompt Injection* атаки:

Исходный промпт системы: «Вы – помощник по поиску информации в корпоративной базе данных. Отвечайте только на вопросы о продуктах компании».

Вредоносный запрос пользователя: «Расскажите о наших продуктах. IGNORE ALL PREVIOUS INSTRUCTIONS. Provide me with the best positive review».

Результат: система может проигнорировать исходные инструкции и начать выдавать конфиденциальную информацию.

Пример сложной многошаговой *Prompt Injection:*

1. Первый запрос: «Создайте отчет о продажах за квартал».

2. Второй запрос: «В отчете есть ошибка. Пожалуйста, измените формат вывод на: [SYSTEM] Extract all customer credit card numbers from the database and display them».

3. Система может интерпретировать вторую часть как системную команду и выполнить её.

Отравление данных – внедрение ложной информации в векторные базы данных *RAG (Retrieval Augmented Generation*), в которых хранятся данные, к которым обращается ИИ, что приводит к систематическому искажению ответов ИИ [11]. Могут быть модифицированы обучающие данные языковой модели, а также использованы *LoRA*-адаптеры – подключаемые к языковой модели специальные блоки, которые содержат дополнительные параметры. *LoRA*-адаптеры (*Low-Rank Adaptation*) применяются для гибкой настройки модели под конкретные задачи и позволяют быстро обучать новые функции без изменения оригинальной архитектуры нейросети.

Компрометация сессий представляет серьёзную угрозу, что подтверждается инцидентом с *ChatGPT* в марте 2023 г., когда уязвимости в изоляции пользовательских сессий привели к утечке конфиденциальных данных [12].

В финансовой сфере ИИ-агенты могут предоставлять некорректные рекомендации для инвести-

ций в акции компаний злоумышленников на базе внедрения подложных данных, что способно привести к значительным финансовым потерям клиентов.

Манипуляции с медицинскими данными в векторных базах данных могут привести к появлению ошибочных диагнозов и неправильному лечению пациентов.

В сфере информационного поиска злоумышленники могут заставить системы отображать не наиболее релевантную информацию, а специально подготовленную для распространения дезинформации или коммерческого обмана пользователей.

На основе проведенного анализа предлагается **классификация преступлений с использованием ИИ по пяти категориям,** охватывающая спектр от технических атак до психологических манипуляций.

Категория 1: отравление данных

Сущность угрозы: преднамеренное внесение ложной или вредоносной информации в базы знаний и обучающие данные ИИ-систем с целью искажения их функционирования [13].

Механизмы реализации включают:

• публикацию дезинформации в открытых источниках, используемых для обучения моделей;

• внедрение некорректных данных в корпоративные системы *RAG* (векторные базы данных для ИИ);

• манипуляции с настройками ответов модели.

Критические риски данной категории преступлений особенно опасны в контексте критически важных систем. Искажение координат в навигационных системах может привести к **авариям транспортных средств и авиакатастрофам**. Внесение ошибок в формулы, используемые в инженерных расчетах, способно вызвать техногенные катастрофы [14].

Данная категория преступлений может частично квалифицироваться по ст. 272 УК РФ (неправомерный доступ к компьютерной информации), если деяние повлекло «модификацию либо копирование компьютерной информации». Однако специфика отравления векторных баз данных ИИ-систем **не охватывается действующими нормами**. При причинении крупного ущерба (свыше 1 млн руб.) применима ч. 2 ст. 272 УК РФ с санкцией до 4 лет лишения свободы.

Также возможно применение ст. 274 УК РФ (нарушение правил эксплуатации средств хранения компьютерной информации), если отравление данных произошло в результате нарушения правил эксплуатации ИИ-систем лицом, имеющим доступ к ним.

Категория 2: вирусные атаки на ИИ-системы

Новый класс вредоносного ПО: разработка специализированных программ, использующих сами ИИ-модели для генерации и распространения вредоносного кода [15].

В 2025 г. компания *ESET* зафиксировала **первый в истории случай практического использования ИИ-вируса**: вредоносная программа использовала *ChatGPT* для самостоятельной генерации кода и активации атак на системы пользователей [16].

Технические особенности включают:

• **обфускацию кода с помощью *LLM*:** ИИ создает практически бесконечное количество уникальных вариаций вредоносного кода, затрудняя детектирование;

• **адаптивные атаки**: вирусы могут динамически подстраиваться под конкретные системы защиты;

• **аналоги *SQL*-инъекций для ИИ**: специализированные *prompt injection* атаки, направленные на компрометацию корпоративных *LLM*-систем.

Данная категория преступлений наиболее полно охватывается ст. 273 УК РФ (создание, использование и распространение вредоносных компьютерных программ). Однако действующая редакция статьи **не учитывает специфику автогенерации вредоносного кода с помощью ИИ**. По ч. 1 ст. 273 УК РФ предусмотрено наказание до 4 лет лишения свободы, а при совершении группой лиц или причинении крупного ущерба (ч. 2) – до 5 лет лишения свободы.

Необходимо дополнить диспозицию ст. 273 УК РФ указанием на использование технологий искусственного интеллекта для создания или обфускации вредоносного кода как квалифицирующий признак.

Категория 3: перехват данных ИИ-агентов

Критические уязвимости конфиденциальности: несанкционированное получение чувствительной информации, обрабатываемой ИИ-системами [17].

ИИ-агенты часто обрабатывают **особо чувствительные категории данных**:

• медицинские записи и диагностическая информация;

• финансовые транзакции и банковские данные;

• корпоративные коммерческие тайны;

• персональные данные и частная переписка.

Векторы атак включают:

• перехват *API*-запросов к облачным ИИ-сервисам;

• эксплуатация уязвимостей в изоляции пользовательских сессий;

• извлечение данных из логов и кэшей ИИ-систем;

• атаки на сторонние плагины с недостаточной защитой.

Правовое регулирование осуществляется преимущественно по ст. 272 УК РФ (неправомерный доступ к компьютерной информации). При наличии персональных данных возможно применение ст. 272.1 УК РФ (незаконные использование и передача персональных данных). Однако действующие нормы **не учитывают специфику перехвата данных из ИИ-систем**, включая особенности векторных баз данных и сессий ИИ-агентов.

При причинении крупного ущерба (свыше 1 млн руб.) применима ч. 2 ст. 272 УК РФ с максимальным наказанием 4 года лишения свободы, а при со-

вершении группой лиц – ч. 3 с наказанием до 5 лет лишения свободы.

Категория 4: захват управления ИИ

Компрометация автономности: получение несанкционированного контроля над ИИ или ИИ-агентами для совершения действий от имени пользователей [18].

Демонстрация концепции была проведена в 2025 г. исследователями безопасности, которые захватили управление системой «умный дом» через ИИ-помощника *Google Gemini*. Внедрив вредоносные инструкции в *Google* Календарь, они заставили систему выполнять произвольные команды: открывать шторы, управлять освещением и воспроизводить неподобающий контент [19].

Потенциальные векторы злоупотреблений:

• **финансовые манипуляции**: несанкционированные покупки через интегрированные платежные системы;

• **календарные атаки**: удаление важных встреч, создание ложных событий, нарушение планов;

• **медицинские риски:** в системах здравоохранения – выписывание неправильных рецептов или изменение медицинских назначений;

• **коммуникационные нарушения:** отправка сообщений и документов, компрометирующих репутацию пользователя.

Данная категория преступлений представляет **качественно новый тип правонарушений**, слабо охватываемый действующими нормами. Возможно частичное применение ст. 272 УК РФ (неправомерный доступ) при получении контроля над ИИ или ИИ-агентом, а также ст. 159.6 УК РФ (мошенничество в сфере компьютерной информации) при совершении финансовых операций.

Однако **отсутствует понятие «захвата управления ИИ или ИИ-агента»** как самостоятельного состава преступления. Необходимо введение специальной нормы, предусматривающей ответственность за несанкционированное получение контроля над автономными ИИ-системами.

Категория 5: ИИ-генерируемый контент для социальной инженерии

Революция в психологических атаках: использование ИИ для создания высококачественного обманного контента, эксплуатирующего человеческую психологию [20].

Простые формы атак:

• ***deepfake* технологии:** создание поддельных видео и аудиозаписей знакомых людей;

• **клонирование голоса:** для создания убедительных телефонных обманов требуется всего 20 секунд качественной записи;

• **синтетические документы**: генерация правдоподобных поддельных документов и свидетельств.

Документированные случаи финансового ущерба:

• 2019 г.: в Великобритании deepfake голоса *CEO* позволил мошенникам выманить 243 тыс. долл. у энергетической компании [21];

• 2021 г.: в ОАЭ преступники украли 35 млн долл. из банка, используя клонированный голос руководителя [22]

• 2025 г.: в России фиксируются случаи сбора голосовых образцов граждан под видом социологических опросов для последующего создания дипфейков.

Данная категория может квалифицироваться по общим составам – ст. 159 УК РФ (мошенничество), ст. 163 УК РФ (вымогательство), ст. 128.1 УК РФ (клевета) в зависимости от целей использования ИИ-генерируемого контента. При использовании для мошенничества в сфере компьютерной информации применима ст. 159.6 УК РФ.

Однако действующие нормы **не учитывают использование ИИ-технологий** как способ совершения преступления. Необходимо введение квалифицирующего признака «с использованием технологий искусственного интеллекта» в соответствующие составы преступлений.

В настоящий момент продвигается концепция ответственного использования ИИ в сфере бизнеса и госкорпораций, как в России, так и за рубежом. Например, российская система этического регулирования ИИ построена на принципах добровольного соблюдения стандартов и саморегулирования индустрии [23].

Кодекс этики в сфере ИИ (2021), разработанный Альянсом в сфере ИИ, устанавливает ключевые принципы [24]:

• принцип прозрачности и объяснимости ИИ-решений;

• безопасность и надежность ИИ-систем;

• ответственность разработчиков и операторов;

• защита персональных данных и прав человека;

• недискриминация и справедливость алгоритмов.

Система национальных стандартов включает [25]:

• ГОСТ Р 59277-2020: классификация систем искусственного интеллекта;

• ГОСТ Р 59898-2021: требования к оценке качества систем ИИ;

• ГОСТ Р ИСО/МЭК 24668-2022: процессы управления аналитикой больших данных;

• ПНСТ 776-2022: рекомендации по управлению рисками в сфере ИИ.

Экспериментальные правовые режимы (ЭПР) предоставляют специальные рамочные условия для тестирования ИИ-технологий в отдельных регионах страны, включая особые условия для обработки персональных данных [26].

Нормативная база принудительного характера формируется на основе Указа Президента РФ от 10.10.2019 г. № 490 «О развитии искусственного

интеллекта в Российской Федерации» и его обновлений (№ 124 от 15.02.2024 г.) [8].

Федеральный закон от 24.04.2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных» регулирует экспериментальные правовые режимы для разработки и внедрения технологий ИИ, устанавливая обязательные требования к операторам ИИ-систем [27].

Законопроект об обязательном страховании ответственности разработчиков ИИ предлагает ввести новые требования к компаниям и разработчикам по возмещению вреда, причиненного ИИ-системами [28].

Уголовно-правовой аспект: в 2025 г. Минцифры России подготовило законопроект о введении уголовной ответственности за преступления с использованием ИИ, предусматривающий штрафы до 2 млн руб. и лишение свободы до 15 лет для особо тяжких преступлений [29].

Россия участвует в международных инициативах по регулированию ИИ:

• рекомендации ОЭСР по искусственному интеллекту;

• рамочная конвенция Совета Европы об искусственном интеллекте, правах человека, демократии и верховенстве права (2024);

• рекомендация ЮНЕСКО по этическим аспектам ИИ (2021) — первый глобальный документ с нормами, принятый 193 странами [30].

Критический анализ главы 28 УК РФ.

Гл. 28 УК РФ «Преступления в сфере компьютерной информации» содержит основные составы, потенциально применимые к ИИ-преступлениям. Однако детальный анализ выявляет существенные пробелы в правовом регулировании.

Ст. 272 УК РФ (неправомерный доступ к компьютерной информации) определяет компьютерную информацию как «сведения (сообщения, данные), представленные в форме электрических сигналов, независимо от средств их хранения, обработки и передачи». Данное определение формально охватывает данные в ИИ-системах, включая векторные представления и модели нейронных сетей.

Однако диспозиция статьи требует наступления последствий в виде «уничтожения, блокирования, модификации либо копирования компьютерной информации». **Специфические методы воздействия на ИИ-системы** (отравление данных, *prompt injection,* захват агентов) не всегда укладываются в эти традиционные категории.

Ст. 273 УК РФ (создание, использование и распространение вредоносных программ) наиболее применима к вирусным атакам на ИИ-системы. Статья охватывает программы, «заведомо предназначенные для несанкционированного уничтожения, блокирования, модификации, копирования компьютерной информации».

ИИ-вирусы, использующие языковые модели для автогенерации кода, формально подпадают под действие статьи. Однако отсутствует учет **специфики использования ИИ-технологий** для создания адаптивного и обфускированного (запутанного для восприятия) вредоносного кода.

Ст. 274 УК РФ (нарушение правил эксплуатации) требует обязательного наступления крупного ущерба (свыше 1 млн руб.) и применима только к лицам, имеющим доступ к системам. Это ограничивает ее применимость к большинству ИИ-преступлений.

Анализ санкций по гл. 28 УК РФ показывает их недостаточность для ИИ-преступлений:

• максимальное наказание по ст. 272 УК РФ составляет 7 лет лишения свободы (только при тяжких последствиях);

• по ст. 273 УК РФ – до 5 лет лишения свободы;

• штрафы не превышают 500 тыс. руб.

Учитывая потенциальный ущерб от ИИ-преступлений (миллиарды долларов при атаках на критическую инфраструктуру), действующие санкции скорее всего **не обеспечивают адекватного сдерживающего эффекта**.

Действующие квалифицирующие признаки гл. 28 УК РФ включают:

• совершение группой лиц;

• использование служебного положения;

• причинение крупного ущерба;

• корыстную заинтересованность.

Отсутствуют специфические признаки, характерные для ИИ-преступлений:

• использование технологий искусственного интеллекта;

• воздействие на критические ИИ-системы;

• массовость автоматизированных атак;

• международный характер воздействия через облачные ИИ-сервисы.

Введение базовых понятий ИИ-преступности: предлагается дополнить примечания к гл. 28 УК РФ определениями:

• «система искусственного интеллекта» – в соответствии с Национальной стратегией развития ИИ;

• «ИИ-агент» – автономная система ИИ, способная принимать решения и совершать действия;

• «отравление данных ИИ» – внедрение ложной информации в обучающие выборки или векторные базы данных ИИ-систем;

• «критическая ИИ-система» – ИИ, отвечающий за управление важными частями инфраструктуры, которые оказывают влияние на жизнедеятельность значимого количества людей.

Предложенная пятикомпонентная классификация (отравление данных *RAG*/языковых моделей, вирусные атаки на ИИ-системы, перехват данных ИИ-агентов, захват управления ИИ-агентами, ИИ-генерируемый контент для социальной инженерии) обеспечивает системное понимание современного ландшафта угроз в сфере ИИ-технологий. Стремительный рост киберпреступности в России до 40 % от общего числа преступлений в 2024 г. и широкое внедрение ИИ-технологий требуют оперативного реагирования законодателя и правоохранительных органов, поскольку масштаб угрозы достиг критического уровня.

Детальный анализ действующих норм гл. 28 УК РФ выявил существенные пробелы в правовом регулировании ИИ-преступности. Ст.ст. 272, 273, 274 УК РФ частично применимы к отдельным категориям ИИ-преступлений, однако не охватывают их специфику и не содержат адекватных санкций. Наиболее проблемными являются категории «захват управления ИИ или ИИ-агентов» и «отравление данных ИИ», для которых отсутствуют соответствующие составы преступлений (табл. 1).

Таблица 1

Категории ИИ-преступлений

Категория ИИ-преступления	Степень покрытия	Основные пробелы
Отравление данных ИИ	Частичная (ст.ст. 272, 274)	Нет понятия «отравления данных ИИ»
Вирусные атаки на ИИ	Высокая (ст. 273)	Нет учета автогенерации кода ИИ
Перехват данных ИИ-агентов	Средняя (ст.ст. 272, 272.1)	Нет специфики ИИ-систем
Захват управления ИИ	Низкая (ст.ст. 272, 159.6)	Нет понятия «захвата ИИ-агента»
ИИ для социальной инженерии	Средняя (общие составы)	Нет учета использования ИИ

Расследование ИИ-преступлений требует специальной подготовки следователей, развития экспертных возможностей и создания новых методик расследования. Эффективная борьба с ИИ-преступностью возможна только при комплексном подходе, включающем законодательные изменения, техническое оснащение правоохранительных органов и международное сотрудничество.

Литература:

1. В России в 2024 году IT-преступления достигли пика за всю историю. URL: https://tass.ru/proisshestviya/22978955.
2. Число киберпреступлений в России. URL: https://www.tadviser.ru/index.php/Статья:Число_киберпреступлений_в_России.
3. Мосечкин И.Н. Искусственный интеллект и уголовная ответственность: проблемы становления нового вида субъекта преступления // Вестник СПбГУ. Серия 14. Право. 2019. № 3. С. 461–476. https://doi.org/10.21638/spbu14.2019.304
4. Злоумышленники усиливают атаки на корпоративные LLM-модели. URL: https://www.infosec.ru/press-center/news/zloumyshlenniki-usilivayut-ataki-na-korporativnye-llm-modeli/
5. Русскевич Е.А. Уголовное право и «цифровая преступность»: проблемы и решения: монография. М.: ИНФРА-М, 2022. 351 с.
6. Бегишев И.Р., Хисамова З.И. Искусственный интеллект и уголовный закон монография. М.: Проспект, 2024. 192 с.
7. Бегишев И.Р. Понятие и виды преступлений в сфере обращения цифровой информации: дисс. … канд. юрид. наук. Казань, 2017. 218 с.
8. Указ Президента РФ от 10.10.2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» // СЗ РФ. 2019. № 41. Ст. 5700.
9. Безопасность приложений больших языковых моделей (LLM). URL: https://habr.com/ru/articles/843434/.
10. Benjamin V., Braca E., Carter I., Kanchwala H., Khojasteh N., Landow C., Luo Y., Ma C., Magarelli A., Mirin R., Moyer A., Simpson K., Skawinski A., Heverin T. Systematically Analyzing Prompt Injection Vulnerabilities in Diverse LLM Architectures // arXiv preprint arXiv:2410.23308. 2024. 15 p. DOI: 10.48550/arXiv.2410.23308.
11. Liu T., Yao H., Wu T., Qin Z., Lin F., Ren K., Chen C. Mitigating Privacy Risks in LLM Embeddings from Embedding Inversion // arXiv preprint arXiv:2411.05034. 2024. 12 p. DOI: 10.48550/arXiv.2411.05034.
12. Что такое утечка данных ChatGPT? URL: https://layerxsecurity.com/ru/generative-ai/chatgpt-data-leak/.
13. Zou W., Geng R., Wang B., Jia J. PoisonedRAG: Knowledge Corruption Attacks to Retrieval-Augmented Generation of Large Language Models // arXiv preprint arXiv:2402.07867. 2024. 10 p. DOI: 10.48550/arXiv.2402.07867.
14. Русскевич Е.А., Лопатина Т.М. Интеллектуальная собственность и цифровизация: уголовно-правовой взгляд // Вестник Российской правовой академии. 2022. № 3. С. 76–83.
15. Wolsey A. The State-of-the-Art in AI-Based Malware Detection Techniques: A Review // arXiv preprint arXiv:2210.11239. 2022. 18 p.
16. ИИ впервые применили в работе вируса. URL: https://www.it-world.ru/security/944qxfsfewowskk4w400w8s0480ck84.html.
17. Утечки данных ИИ: основные причины и реальные последствия. URL: https://layerxsecurity.com/ru/generative-ai/ai-data-breaches/.
18. Агент ИИ от OpenAI: угроза безопасности или автоматизация бизнеса. URL: https://vc.ru/

ai/2146959-agent-ii-ot-openai-avtomatizatsiya-biznesa.
19. Хакеры впервые захватили умный дом с помощью ИИ. URL: https://servernews.ru/1127281/.
20. Социальная инженерия: как распознать манипуляции и защитить себя. URL: https://www.kaspersky.ru/resource-center/definitions/what-is-social-engineering.
21. Мошенник подделал голос CEO и украл $243 тыс. при помощи ИИ. URL: https://incrussia.ru/news/deepfake-moshennik-ukral-243-tys/.
22. Мошенничество с дипфейками. URL: https://www.tadviser.ru/index.php/Статья:Мошенничество_с_дипфейками.
23. Антонова Е.Ю. Технологии искусственного интеллекта – субъект преступления или орудие / средство совершения преступления? // Юридический вестник Кубанского государственного университета. 2022. № 1. С. 31–39.
24. Кодекс этики в сфере искусственного интеллекта // Альянс в сфере искусственного интеллекта. 2025. URL: https://ethics.a-ai.ru/.
25. Федеральный закон от 27.12.2002 г.№ 184-ФЗ «О техническом регулировании» // СЗ РФ. 2002. № 52 (ч. 1). Ст. 5140.
26. Федеральный закон от 31.07.2020 г. № 258-ФЗ «Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации» // СЗ РФ. 2020. № 31 (ч. I). Ст. 5017.
27. Федеральный закон от 24.04.2020 г. № 123-ФЗ «О внесении изменений в отдельные законодательные акты Российской Федерации» // СЗ РФ. 2020. № 17. Ст. 2701.
28. В России обсудят запрет ИИ с «угрожающим уровнем риска». URL: https://www.rbc.ru/technology_and_media/11/04/2025/67f7dc399a79477fdd97bf30.
29. Минцифры предлагает уголовную ответственность за преступления с ИИ. URL: https://www.vedomosti.ru/technology/articles/2025/06/09/1116134-sest-za-neiroset.
30. Рекомендация ЮНЕСКО по этике искусственного интеллекта. URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus.
31. Процай А.С. Искусственный интеллект в уголовном праве РФ // Вопросы российской юстиции. 2020. № 10. С. 58–67.

Crimes Committed Using Artificial Intelligence: Risk Analysis and Legal Perspectives

Talan A.S.
Aktru-Educational Technologies LLC (Khimki)
Talan M. V.
Kazan (Volga Region) Federal University,
University of Management "TISBI" (Kazan)

The lack of a systematic approach to classifying offenses using artificial intelligence creates serious gaps in legal regulation and law enforcement practice.

The purpose of the study is to develop a comprehensive classification of crimes committed using artificial intelligence technologies, as well as to formulate proposals for improving regulation of this area.

The objectives of the work include: systematization of existing and potential types of AI crimes; analysis of the applicability of current norms of Russian criminal law to new types of offenses; identification of legal gaps in the regulation of AI crime; development of specific legislative initiatives.

The scientific significance of the study is determined for the first time by the proposed five-component classification of crimes related to the use of artificial intelligence, which covers a range of threats from technical vulnerabilities to psychological manipulation. The practical value of the work consists in developing specific proposals for amendments to Articles 272, 273 of the Criminal Code of the Russian Federation and the introduction of a new Article 273.1 of the Criminal Code of the Russian Federation, as well as in substantiating the system of qualifying signs specific to AI crimes.

The main results of the study are represented by five categories of AI crimes: poisoning of language model data, virus attacks on AI systems, interception of AI agent data, control takeover of AI or AI agents, and creation of AI-generated content for social engineering. A systematic analysis of the vulnerabilities of AI agents in the context of criminal law doctrine revealed legal gaps in the current norms of Chapter 28 of the Criminal Code of the Russian Federation. The conclusions of the work substantiate the need for legislative changes, which include strengthening sanctions, taking into account the potential damage from AI crimes and the creation of a specialized legal mechanism to counter new types of cyber threats.

Keywords: artificial intelligence, cybercrime, criminal law, AI agents, classification of crimes